数字人文视角下的古籍知识关联探析

王丽丽[1,2,3]，张　宁[1,2*]

（1.中国人民大学 图书馆，北京　100872；2.中国人民大学 古籍整理研究所，北京　100872；

3.中国人民大学 数字人文研究中心，北京　100872）

摘　要：［目的 / 意义］数字人文方兴未艾，为古籍知识关联的实现提供了重要现代科技手段。对古籍内容挖掘与知识关联进行研究，有助于提升古籍资源开发水平与服务质量，推动中华优秀传统文化创造性转化、创新性发展。［方法 / 过程］通过文献综述和网络调研，梳理数字人文视角下古籍知识关联的研究，探讨古籍知识关联的关键环节，通过古籍知识关联应用举例分析古籍知识关联的壁垒。［结果 / 结论］数字人文视角下，从文献单元层次的组织深入到知识单元层次的组织，借助数据分析技术、机器学习、可视化技术等，可实现古籍知识关联。古籍知识关联起点是文献组织，基础是古籍数据，本质是知识组织。古籍知识关联可进行方志与家谱挖掘、钤印知识网络发现、特定时期与地域学者研究脉络与交往分析等，但在实施中仍面临知识元处理难度大、技术要求高、集成平台少等壁垒。

关键词：古籍整理；数字人文；知识关联；知识元

中图分类号：G250　　　**文献标识码：**A　　　**文章编号：**1002-1248（2022）09-0051-09

引用本文：王丽丽，张宁. 数字人文视角下的古籍知识关联探析［J］. 农业图书情报学报，2022，34（9）：51-59.

1 引　言

党的十八大以来，以习近平同志为核心的党中央对传承和弘扬中华优秀传统文化作出一系列重大决策部署，古籍事业迎来新的发展机遇。2022 年 3 月，“加强文物古籍保护利用”首次写入政府工作报告[1]。4 月 11 日，中共中央办公厅、国务院办公厅印发《关于推进新时代古籍工作的意见》[2]。4 月 25 日，习近平总书记在考察中国人民大学图书馆时指出，要运用现代科技手段加强古籍典藏的保护修复和综合利用[3]。5 月 27 日，习近平总书记强调：把考古探索和文献研究同自然科学技术手段有机结合起来[4]。

古籍具有独特的文物价值、学术价值和艺术价值，是凝聚传统文化最重要的载体之一。从传统的古籍文献书目整理，到古籍书目和全文数据库建设，对古籍资源的开发与利用一直是学界和业界关注的焦点。当前，古籍资源的数字化著录已较为普及，如各单位自建古籍书目数据库、“学苑汲古：高校古文献资源库”“全国古籍普查登记基本数据库”等。此外，还出现一批商业化的古籍全文数据库，如“中国基本古籍库”“中华经典古籍库”“书同文古籍数据库”等。但随着

收稿日期：2022-06-07

基金项目：2021—2023 年北京高校图书馆研究基金项目“数字人文环境下红色文献共建共享与挖掘利用探析”（BGT2021040）

作者简介：王丽丽，女，博士，副研究馆员，中国人民大学图书馆，研究方向为知识组织、特色资源整理

*通信作者：张宁，女，硕士，中国人民大学图书馆，研究方向为特色资源整理。Email：zznn@ruc.edu.cn

大规模古籍数据库的建成，学者发现古籍数据库只是对古籍文献外部特征与内容特征进行揭示，仅能满足用户基本的检索与使用的需求，并未对古籍资源从知识组织的角度进行深入挖掘和利用，传统古籍整理的局限性逐渐暴露出来[5]。

近年来，计算机信息技术与人文学科研究深度结合的新兴交叉学术领域——数字人文（Digital Humanity）得到了迅速发展，成为加强古籍典藏保护修复和综合利用的重要现代科技手段。数字人文背景下，对古籍资源的整理与研究从数字化向数据化迈进，从全文数据库向知识平台迈进，从全文检索向智慧检索迈进。古籍知识关联指的是汇集大量结构化的古籍数据，对其进行细粒度、多维度切分，通过机器标引、辅以人工标引，利用关联数据、语义网等技术，进行统计分析、网络分析和空间分析等，探寻数据间的关系，实现古籍资源外部与内容特征、古籍资源中蕴含的知识元之间的互联，并可根据用户的需求，实时、自动实现知识之间的自动重组与可视化展示。本文在梳理数字人文视角下古籍知识关联的研究基础之上，探讨古籍知识关联的关键环节，举例分析古籍知识关联，总结古籍知识关联的壁垒。

2 古籍知识关联的研究现状

数字人文的出现为古籍整理带来获取资料的智能化、研究方法的数据化、研究结果的可视化三大进展[6]。将数字人文应用于古籍整理，改变了古籍文献传统检索方式，从电子文献到结构化数据库、从点状检索到网状关联、从逐条拷贝到分类打包，使传统的静态文本转化为可随意组合的动态文本[7]。古籍整理不再仅仅是实现更大范围的文献信息的揭示与保存，更重要的是要实现古籍文献蕴含知识之间的关联，深入挖掘其蕴含的哲学思想、人文精神、价值理念、道德规范。知识与知识之间互相发生联系与影响即为知识关联。对知识关联的研究已经有不少，如文庭孝[8,9]围绕知识关联的基础理论、内部特征、结构分析与应用开展研究，赵蓉英研究了[10]知识关联的类型。也有学者对学术文献及其知识关联[11]、金融领域知识关联[12]等特定领域文献间的知识关联进行研究。目前围绕古籍资源知识关联进行的理论研究主要是围绕某一数字人文平台展开的。李兵等对中医古籍知识化研究现状和常用的知识分析与挖掘方法进行综述[13]。黄水清认为应利用数字人文技术实现古籍资源的创造性转化与创新性发展。欧阳剑提出在新时代古籍资源转化方面，文本化和数据化是基础、知识化是重要方式、平台化是有效利用及传播的重要途径[14]。魏晓萍探讨了数字人文理论与技术方法在古籍文献开发与利用的应用[5]。李欣等认为数字方志特藏资源建设应使用关联书目元数据方案，最小粒度化方志元数据，提供开放平台，元数据的创建、分析等可以利用文本分析技术等技术手段实现或者以众包形式开展[15]。王蕾等从元数据方案、检索点设计和数据关联等角度，总结徽州文书数据库的建设经验[16]。陈力认为数字人文视域下的古籍数字化与古典知识库建设存在计算机编码汉字和计算机图像识别并转换为字符的能力等问题[17]。

理论研究指导实践发展，利用自然语言处理技术、关联数据、GIS 技术、可视化技术等，数字人文指导下古籍知识关联领域已经产生了较为丰硕的实践成果。一是通过自动标注、自动校对、词语切分等自然语言处理技术，实现命名实体的识别与关联等。如清华大学自然语言处理与社会人文计算实验室研发的中文诗歌自动生成系统《九歌》，以大规模诗歌文本为研究对象分析诗歌内在联系，为用户提供诗歌生成以及人机交互创作等功能[18]；侯汉清、包平等围绕《方志物产》中地名等命名实体的识别与关联开展了系列研究[19]。二是通过关联数据，实现古籍文献中的时间、地点、人物等的关联，形成知识网络，挖掘古籍文献的内在关系。如“中国历代人物传记资料库（CBDB）”构建关系型数据库为历史文献人物建模[20]；上海图书馆“中国家谱知识服务平台”应用关联数据重构上海图书馆的家谱服务。三是结合 GIS 技术呈现古籍文献资源的时空分布。如“学术地图发布平台”实现了数据共享、可视化分析及多功能查询等，“唐宋文学编年地图平台”以编年地图的形式提供唐宋时期文学人物、

事件的时空分布及作品关系。四是通过可视化技术等构建知识图谱。如北京大学“宋元学案知识图谱可视化系统”梳理宋元时期学术史中的关系图谱、学术流变、师承关系、学派传承，构建完整的学术史衍化脉络全景；武汉大学数字文化遗产研究中心完成了“数字敦煌莫高窟”“新疆克孜尔石窟数字化”“颐和园佛香阁三维重建”等项目。

相较于传统意义上的古籍数据库，这些实践成果有了新的发展。一是组织机制不同，传统的古籍整理目的在于版本鉴定、揭示资源、保存资源、利用资源，而古籍数字人文平台功能更为多样、深入，其融合馆藏元数据记录、专家研究成果、相关网络资源，以知识本体为基础，应用关联数据等技术，对古籍资源完成数字化到数据化的知识重组，以可视化的形式揭示古籍资源间的知识关联。二是收录范围扩展，传统的古籍数据库收录范围主要局限在原始文献资料方面[21]，古籍数字人文平台不仅收录原始文献，还收录古籍文献的解析文本、延伸文献等，更在一定程度上实现了从“数字化”到“数据化”的过渡，从“数据孤岛”到“数据共享”的转变。以往的数据库平台多为“数字化”平台，主要为全文扫描图像与文本存储平台；而古籍数字人文平台则是一个“数据化”的平台，从“存储”提升到“分析”与“关联”。

3 古籍知识关联的关键环节

纵览这些古籍数字人文平台，古籍知识关联是在古籍文献组织基础之上，基于古籍文本数据进行的知识单元间的关联。

3.1 古籍知识关联的起点——文献组织

在原有的古籍数据库基础之上，数字人文视角下的古籍知识关联可以实现更大范围文献间的组织。

（1）平台文献之间的组织。从揭示程度来看，数字人文视角下的古籍平台不仅仅实现了书目信息的集合揭示，还实现了对书目信息的区分、聚类、比较和统计分析[22]，进一步对文本化的古籍内容进行数据化转换。一方面，借助于相似的数字化技术和统一的元数据标准，分散、独立的各古籍平台间的书目数据可进行互联，实现不同平台间的关联；另一方面，在平台内部，由于古籍数字人文平台对资源的内外部特征描述更为全面，可实现繁简字、异体字、版本之间(汇编本与单行本、丛书的著录）等的互联。

（2）平台内外文献之间的组织。这主要指的是平台文献与新印古籍资源、延伸研究文献、互联网学术资源等之间的关联与组织。对于人文学者来说，古籍原本的阐释是其研究工作的基础，依据这些原始文献开展的相关研究也是重要的研究资料。新印古籍资源不仅仅是原典的简单影印出版，更是专题性、纵深性资料的精选汇总。延伸研究文献指的是和某一古籍相关的新出版的学术研究论著、学者年表、编年史等研究资料。如家谱资源的延伸研究文献包括人物传记、历史年表、职官表、日记信札、报纸公告栏等。将中国知网、超星等数据库中相关的研究资料关联到古籍数字人文平台之上，将大大拓展用户学术研究的视野，节约其时间。如清代台湾方志物产篇分析系统利用学名将方志中的物产联结到现代植物资料库，将西方分类学观念建构的知识体系与传统方志的资料互相联结[23]。

3.2 古籍知识关联的基础——古籍数据

文献组织是对古籍书目信息的组织，知识组织是对古籍文本的组织。而古籍文本不能直接被计算机语言识别和使用，需要将其转化为可识别、可理解的结构化数据，进一步从数据转化为知识，方能实现古籍知识关联。因此，古籍数据是古籍知识关联的基础。

（1）古籍文本转化为古籍数据。首先，清洗以图片、文本等格式存储的古籍数据，通过 OCR 技术等采集古籍文本数据，借助分词与标注技术等将古籍数据颗粒化，形成具有独立意义的结构化数据单元，借助大数据技术对各数据单元进行标引形成新的聚类，以实现数据的重新组织。如中文古籍半自动化标记平台(MARKUS）通过关联多个权威语料库实现古籍中历史人名、地名、官名与时间等实体的自动标注，为研究者提供可以通过定义关键字列表、上下文中的关键字、

正则表达式等方式进行半自动标注，且支持以 txt、excel、html 格式输出标注结果[24]。

（2）古籍数据的量化分析，包括统计分析、网络分析和空间分析。古籍数据的统计分析主要是利用数学语言对独立数据单元的总数、频次等进行概括性分析，较易实现。古籍数据的网络分析是指对已颗粒化的古籍数据以某一共同特征构建出新网络结构进行分析，如《宋元学案》知识图谱可视化系统将《宋元学案》人物、时间、地点、著作及它们之间的复杂语义关系提取出来构造成知识图谱，提供可视化展现、交互式浏览、语义化查询等功能[25]。古籍数据的空间分析即是将线性平面的数据实现点线面的空间分布，包括结构图、地图、分布图等，如“学术地图发布平台”借助 GIS 技术实现历史人物的行迹、特征分布和社会关系等信息的地理映射。

3.3 古籍知识关联的本质——知识组织

以文献为单元的传统古籍数据库仅能满足用户的资料需求，远不能满足其知识需求，古籍中所包含的细粒度知识元的组织成为古籍数字人文平台的研究重点，这也是古籍知识关联的本质。知识元的概念最早于 20 世纪 70 年代后期由弗拉基米尔·斯拉麦卡提出，学术文献中的知识元指的是语义上相对完整地表达特定知识的最小的内容单元[26]。古籍典籍知识元的涵义又比普通学术文献中的知识元更为特殊，其类型更为多样、复杂。

不同类别的古籍所包含的知识元类型有所不同。以史部为例，其下属 17 大类中，地理类之下的总志、方志、专志等多包含的知识元类型相似，数字方志集成平台即实现了对方志类知识元的组织。该平台拆分、合并华东师范大学图书馆等图书馆、超星、CADAL、中国方志库、瀚堂典籍库及方正电子书等方志类元数据，采用 BIBFRAME 书目数据模型进行转换、查重、映射，实现方志与外部数据的互联、最小粒度化方志元数据、平台开放、用户可协作参与，并尝试通过内容分析技术对方志中的内容进行分析[15]。史部史表类、传记类的人表、年谱、科举录、职官录知识元类型相似，多表现为“某人生于某年，任某职”等。如《山阴州山吴氏族谱》中有“太学生楚材公，姓吴氏，讳乘权；子立公，讳乘业。山阴州山里人。”包含了“楚材公是太学生”“楚材公姓吴氏”“楚材公讳乘权”“子立公讳乘业”“楚材公和子立公为山阴州山里人”等语义三元组，以 RDF 数据格式描述，将人、地、时串联成一个个的迁徙事件，可将具有共同先祖的先祖名人及其迁徙事件关联起来[27]。而史部金石类中的知识元不仅包括对图像的识别，即如何将图像中的内容进行抽取、释读，还包括对识别后的文本知识元进行语义描述。

除此之外，古籍资源还包含一些特殊类型的知识元。如版本知识元、钤印知识元，通过对“楝亭曹氏藏书”印主为曹寅、“楝亭曹氏藏书”印为长方形、阳文、篆体，“楝亭曹氏藏书”刻于《皇明大训记》书名页等的描述，可以让机器准确理解钤印知识元，如图 1 所示。

4 古籍知识关联的应用举例

从文献单元的组织深入到知识单元的组织，古籍数字人文平台可实现知识关联。关联的本质是实现文献与文献、知识与知识之间的互相联系与影响。古籍数字人文平台实现了更大范围书目数据的聚合，有利于学者将散落在不同机构的同一作品的不同版本聚集，“辨章学术、考镜源流”，进行版本鉴定与版本源流的考证。而借助内容挖掘、时空分析等数据分析技术与知识地图、主题图等可视化技术，古籍知识关联可进行方志与家谱挖掘、钤印知识网络发现、特定时期与地域学者研究脉络与交往分析等。

4.1 方志与家谱挖掘

夫家有谱、州有志、国有史，方志和家谱具有揭示某一区域与家族长期、延续的发展状况与迁徙路线的史料性价值。通过集成散落在各地的大量零散方志、家谱，借助于统一的元数据标准，可集成不同平台间志谱书目数据、志谱相关资源。在此基础之上，对大

RDF数据		
	P	O
http://data.library.sh.cn/gj/entity/seal/tl2ta1su4fj8ks15	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.library.sh.cn/ontology/Seal
http://data.library.sh.cn/gj/entity/seal/tl2ta1su4fj8ks15	http://www.w3.org/2000/01/rdf-schema#label	楝亭曹氏藏书
http://data.library.sh.cn/gj/entity/seal/tl2ta1su4fj8ks15	http://www.w3.org/2000/01/rdf-schema#label	楝亭曹氏藏書
http://data.library.sh.cn/gj/entity/seal/tl2ta1su4fj8ks15	http://www.library.sh.cn/ontology/owner	http://data.library.sh.cn/entity/person/y1bxgqoz33fksx75
http://data.library.sh.cn/gj/entity/seal/tl2ta1su4fj8ks15	http://xmlns.com/foaf/0.1/img	http://data.library.sh.cn/gj/resource/img/i2uzwhkpuilqnfrs
http://data.library.sh.cn/gj/entity/seal/tl2ta1su4fj8ks15	http://www.library.sh.cn/ontology/libliotheca	http://data.library.sh.cn/gj/entity/libliotheca/v8bvwuw1czl6o3rm
http://data.library.sh.cn/gj/entity/seal/tl2ta1su4fj8ks15	http://www.library.sh.cn/ontology/ownerOfSeal	曹寅
http://data.library.sh.cn/gj/entity/seal/tl2ta1su4fj8ks15	http://www.library.sh.cn/ontology/sealCharacters	楝亭曹氏藏书
http://data.library.sh.cn/gj/entity/seal/tl2ta1su4fj8ks15	http://www.library.sh.cn/ontology/sealCharacters	楝亭曹氏藏書
http://data.library.sh.cn/gj/entity/seal/tl2ta1su4fj8ks15	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://xmlns.com/foaf/0.1/Image
http://data.library.sh.cn/gj/entity/seal/tl2ta1su4fj8ks15	http://www.library.sh.cn/ontology/description	楝亭曹氏藏書
http://data.library.sh.cn/gj/entity/seal/tl2ta1su4fj8ks15	http://purl.org/dc/elements/1.1/source	明清藏书家藏书印
http://data.library.sh.cn/gj/entity/seal/tl2ta1su4fj8ks15	http://www.library.sh.cn/ontology/imgAbout	http://data.library.sh.cn/gj/entity/seal/tl2ta1su4fj8ks15

* 数据来源：上海图书馆开放数据平台·印章知识库（http://data1.library.sh.cn/gj/webapi/toSealList）

图 1　“楝亭曹氏藏书”藏书印 RDF 数据

Fig.1　RDF data of " Cao's collection" book seal

规模、多源、多维的志谱资源进行文本化、数据化转换，抽取其中蕴含的知识单元，通过量化分析与可视化展示，实现志谱的数据集成与知识关联。如中国家谱知识服务平台把散落在不同家谱文献中的人、地、时、事关联起来，发现人与人之间、人与文献之间的关系，形成完整的知识图，以可视化的方式展示。这既便于普通读者直观获得隐藏在不同文献中的知识，并将不同的文献按照某一主题有机地组织起来，提供知识导航；更可以帮助研究者发现问题，提出问题[27]。

4.2　钤印知识网络发现

钤印知识元是古籍资源中一种重要的知识元。将钤印作为实体对象，对其所蕴含的印文、印主、藏本知识进行语义标注，实现同一印主不同钤印间的归一处理，实现同一印主的字、号、别名等之间的归一处理，并将藏本信息与联合目录、循证平台关联，将时间、地点与历史纪年表、地理名词表等关联，将印主信息与人名规范档等进行关联，进而发现印文、印主与藏本之间的关联关系，实现平台内外文献的互联与钤印的知识挖掘。通过构建钤印知识元库，既可以加深钤印内含知识的挖掘层次、细化其表示粒度，又可以辅助梳理递藏源流、考证藏书的收藏与散逸情况、溯清馆藏来源[28]。

4.3　特定时期与地域学者研究脉络与交往分析

通过对多源、异构古籍资源的文献组织、知识组织，可实现平台内外的文献与文献之间、文献内外知识单元之间的关联，帮助研究者发现人工难以获得的知识，进行特定时期与特定地域学者研究脉络与交往等的分析。这是当前古籍知识关联实现的重要价值所在。如欧阳剑以大规模中国古籍文本（41 563 种古籍，总计 48.35 亿字）为研究对象，对古籍进行整理、标注、自动分词等处理，创建了一个以语言学、历史文献学、历史地理学等人文学科研究为主的古籍实时统计分析平台，实现了实时统计分析、时间和空间可视化分析，可辅助研究者在大量的古籍文献中发现新的模式、现象、趋势等[29]。

5　古籍知识关联的壁垒

数字人文视角下古籍知识关联虽极具现实意义，但在实施中仍然受到种种壁垒的限制。

5.1 知识关联对象的壁垒

对于大多数机构来说，知识关联的对象——古籍文献数据与知识元的文本化处理与关联化实现仍然具有较高的难度。古字无定形，已完成的古籍数字化资源在文本化转换过程中仍以 OCR 识别加专家人工校读为主要工作方式。虽然目前书同文 i- 慧眼 OCR 平台、古联公司古籍 OCR 系统、如是 OCR 等在古籍文本 OCR 识别方面已经取得了一定的成就，但普遍意义上来说，古籍 OCR 识别始终面临版式多样（横版、竖版、小字双行等）、格式各异（除文字外还有碑帖拓片、舆图等实物、图像类）、异体字众多、字形字体多变、字符集大而标注样本少等困难。

古籍全文数字化是进行古籍知识关联的基础，古籍文本的识别与转换相关研究问题一直深受学界重视。不少学者基于机器学习、深度学习等方法技术，提出了古籍文档图像版面分析方法与文字检测分割方法、构建大规模图像和文本语料库、设计古籍文档图像预训练模型等。未来，应进一步增强该领域的研究，以提高古籍文本识别效率与质量。

5.2 组织方法、技术与工具的壁垒

要实现古籍知识关联，在古籍文本化的基础之上，还需要将传统古籍数据库中的关系型古籍全文数据转换为语义层次的知识元。数据化的古籍知识需要有相关机构牵头，多学科领域专家参与制定统一的、科学的元数据描述规范，方能使得不同平台间文献单元、知识单元实现共享、形成映射。而在知识关联的构建过程中，需要应用到各种技术方法。对于图书馆等古籍资源馆藏机构来说，从文本分析、聚类分类、主题分析、内容挖掘、时序分析、地理空间分析、社会关系分析等内容分析技术到可视化技术，到机器学习的技术、方法需要得到进一步开发与应用，这是发现知识元与知识元间的关系，并以可视化的形式进行揭示，进而实现知识关联的关键。图书馆等馆藏机构应进一步与相关机构加强合作，推动相关技术方法的开发、落地、应用、推广。

5.3 组织体系的壁垒

古籍数字人文项目的纷纷落地、数字人文平台的不断推出也展露了当前古籍数字人文研究的一个隐性弊端，即大大小小的数字人文项目“遍地开花”，大部分古籍数字人文平台仅为某一机构所开发，多针对某一特定类型资源如方志、中医药、佛经等甚至单一部古籍等，使用范围也多受限于建设单位内部。数字人文视角下的古籍知识关联应克服传统古籍整理的“小、散、乱”模式，从建制上进行努力，避免重复建设，参考古籍书目数据库“全国古籍普查登记基本数据库”等模式，由相关机构牵头，馆藏机构、人文学者、计算机领域学者共同协作，多建设如“中文古籍联合目录及循证平台”“数字方志集成平台”这类综合性、集成性、可扩展性的平台，促进古籍资源的共建共享、深度挖掘。

6 结　语

古籍资源是一个图书馆“人无我有”、具有特殊价值的一部分资源，是数字人文研究的重要对象之一，数字人文的理念、工具与方法是古籍资源开发与利用的重要助力。数字人文视角下，从文献单元层次的组织深入到知识单元层次的组织，借助数据分析技术、机器学习、可视化技术等，可实现古籍知识关联。古籍知识关联起点是文献组织，基础是古籍数据，本质是知识组织。古籍知识关联可进行方志与家谱挖掘、钤印知识网络发现、特定时期与地域学者研究脉络与交往分析等，但在实施中仍面临知识元处理难度大、技术要求高、集成平台少等壁垒。

囿于研究条件所限，论文仅从理论角度对古籍知识关联的研究与应用、关键环节、壁垒等进行了论述，未来我们将以馆藏古籍知识平台优化升级为契机，进一步对古籍元数据描述规范、古籍文本转换与内容挖掘、不同平台间数据交换及更新机制进行研究，以促进古籍资源更深层次的知识关联、更大范围的共建共享，提升古籍资源开发水平与服务质量。

参考文献：

[1] 朱羿. 加强古籍保护绵延中华文化[EB/OL]. (2022-03-18)[2022-04-28]. http://ex.cssn.cn/zx/bwyc/202203/t20220318_5399424.shtml.

ZHU Y. Strengthening the protection of ancient books and extending Chinese culture[EB/OL]. (2022-03-18)[2022-04-28]. http://ex.cssn.cn/zx/bwyc/202203/t20220318_5399424.shtml.

[2] 新华社. 中共中央办公厅国务院办公厅印发《关于推进新时代古籍工作的意见》[EB/OL].(2022-04-11)[2022-04-28]. http://www.gov.cn/zhengce/2022-04/11/content_5684555.htm.

Xinhua. China to step up preservation, publication of ancient books[EB/OL].(2022-04-11)[2022-04-28]. http://www.gov.cn/zhengce/2022-04/11/content_5684555.htm.

[3] 习近平. 坚持党的领导 传承红色基因 扎根中国大地 走出一条建设中国特色世界一流大学新路[N]. 人民日报, 2022-04-26(1).

XI J P. Xi calls for blazing new path to develop China's world-class universities[N]. People's daily, 2022-04-26(1).

[4] 人民日报. 习近平: 把中国文明历史研究引向深入 推动增强历史自觉坚定文化自信[N]. 人民日报, 2022-05-29(1).

People's daily. Xi calls for advancing study of Chinese civilization[N]. People's daily, 2022-05-29(1).

[5] 魏晓萍. 数字人文背景下数字化古籍的深度开发利用[J]. 农业图书情报学刊, 2018, 30(9): 106-110.

WEI X P. Deep development and utilization of digital ancient books under the background of digital humanities[J]. Journal of library and information sciences in agriculture, 2018, 30(9): 106-110.

[6] 谢颖, 王兆鹏. 数字化助力学术研究发展——访四川大学文学与新闻学院文科讲席教授王兆鹏[N]. 人民政协报, 2022-06-06(05).

XIE Y, WANG Z P. Digitalization helps academic research development - An interview with Wang Zhaopeng, chair professor of the college of literature and journalism, Sichuan university[N]. CPPCC daily, 2022-06-06(05).

[7] 阮益嫘. 数字化助力古籍活化传承[N]. 中国社会科学报, 2022-05-25(01).

RUAN Y L. Digitalization helps activate and inherit ancient books[N]. Chinese social sciences today, 2022-05-25(01).

[8] 文庭孝, 龚蛟腾, 张蕊, 等. 知识关联: 内涵、特征与类型[J]. 图书馆, 2011(4): 32-35.

WEN T X, GONG J T, ZHANG R, et al. Knowledge connection: Meaning, characteristic and type[J]. Library, 2011(4): 32-35.

[9] 文庭孝, 刘晓英, 刘灿姣, 等. 知识关联的结构分析[J]. 图书馆, 2011(2): 1-7.

WEN T X, LIU X Y, LIU C J, et al. Structure analysis of knowledge connection[J]. Library, 2011(2): 1-7.

[10] 赵蓉英. 论知识网络的结构[J]. 图书情报工作, 2007, 51(9): 6-10.

ZHAO R Y. On the structure of knowledge network [J]. Library and information service, 2007, 51(9): 6-10.

[11] 李睿, 孟连生. 论专利引用行为与期刊论文引用行为在揭示知识关联方面的差异[J]. 情报学报, 2010, 29(3): 474-478.

LI R, MENG L S. Difference between paper citations and patent citations in the presentation of knowledge link[J]. Journal of the China society for scientific and technical information, 2010, 29 (3): 474-478.

[12] 唐旭丽, 马费成, 傅维刚, 等. 知识关联视角下的金融知识表示及风险识别[J]. 情报学报, 2019, 38(3): 286-298.

TANG X L, MA F C, FU W G, et al. Research on financial knowledge representation and risk identification from knowledge connection perspective[J]. Journal of the China society for scientific and technical information, 2019, 38(3): 286-298.

[13] 李兵, 张华敏, 李莎莎, 等. 中医古籍知识深度利用方法与知识库构建[J]. 中国数字医学, 2018, 13(8): 33-35.

LI B, ZHANG H M, LI S S, et al. The deep exploitation of TCM knowledge in ancient books and knowledge base construction [J]. China digital medicine, 2018, 13(8): 33-35.

[14] 黄水清, 王晓光, 夏翠娟, 等. 推进新时代古籍工作, 加快创新智能化发展[J]. 农业图书情报学报, 2022, 34(5):4-20.

HUANG S Q, WANG X G, XIA C J, et al. Advancing the work on ancient classics in the new era and accelerating innovative and intelligent development[J]. Journal of library and information sciences in agriculture, 2022, 34(5): 4-20.

[15] 鲁丹, 李欣. 数字人文环境下异构方志元数据整合策略[J]. 图书馆论坛, 2019(4): 158-165.

LU D, LI X. Research on metadata integration strategy of heterogeneous local chronicles in digital humanities environment[J]. Library tribune, 2019(4): 158-165.

[16] 王蕾, 薛玉, 肖鹏, 等. 民间历史文献数字人文图书馆构建——以徽州文书数字人文图书馆实践反思为例[J]. 图书馆论坛, 2018, 38(3): 30-36.

WANG L, XUE Y, XIAO P, et al. The rise of digital humanities and the development of folk historical literature database: Case study of Huizhou documentation database[J]. Library tribune, 2018, 38(3): 30-36.

[17] 陈力. 数字人文视域下的古籍数字化与古典知识库建设问题[J]. 中国图书馆学报, 2022, 48(2): 36-46.

CHEN L. Digitalization of ancient books and construction of classical knowledge repository from the perspective of digital humanities[J]. Journal of library science in China, 2022, 48(2): 36-46.

[18] 清华大学自然语言处理与社会人文计算实验室. 九歌|自动作诗[EB/OL]. [2022-06-27]. http://nlp.csai.tsinghua.edu.cn/.

Department of computer science and technology, Tsinghua university. Jiu gelautomatic poetry writing[EB/OL].[2022-06-27]. http://nlp.csai.tsinghua.edu.cn/.

[19] 朱锁玲, 包平. 方志类古籍地名识别及系统构建[J]. 中国图书馆学报, 2011, 37(193): 118-124.

ZHU S L, BAO P. System design for location name recognition in ancient local chronicles[J]. Journal of library science in China, 2011, 37(193): 118-124.

[20] 包弼德, 王宏苏, 傅君劢, 等. "中国历代人物传记资料库"(CBDB)的历史、方法与未来[J]. 数字人文研究, 2021, 1(1): 21-33.

PETER K B, WANG H S, MICHAEL A F, et al. The History, Methods, and future of the China biographical database(CBDB) project[J]. Digital humanities research, 2021, 1(1): 21-33.

[21] 王兆鹏. 建设中国文学数字化地图平台的构想[J]. 文学遗产, 2012(2): 131-133.

WANG Z P. The conception of building a digital map platform of Chinese literature[J]. Literary heritage, 2012(2): 131-133.

[22] 夏翠娟, 林海青, 刘炜. 面向循证实践的中文古籍数据模型研究与设计[J]. 中国图书馆学报, 2017, 43(6): 16-34.

XIA C J, LIN H Q, LIU W. Designing a data model of Chinese ancient books for evidence-based practice[J]. Journal of library science in China, 2017, 43(6): 16-34.

[23] 张素玢, 李钰淳. 资料撷取与描述比对——清代台湾方志物产篇分析系统之建置与解释[A]. 数位人文研究与技艺[M]. 台北: 国立台湾大学出版社, 2014: 67-89.

ZHANG S PI, LI Y C. Data extraction and mapping: The taxonomy of fauna and flora of Taiwan local gazetteers in the Qing dynasty[A]. Digital humanities and craft: Technological change[M]. Taipei: National Taiwan university press, 2014: 67-89.

[24] 于亚秀, 李欣. 数字人文视域中的古籍文本标注方法研究——以MARKUS为例[J/OL]. 大数据: 1-11[2022-05-16].http://hfghi60aabc7d15084b00h060wbob5fqnx6n66.fhcz.libproxy.ruc.edu.cn/kcms/detail/10.1321.G2.20220401.1032.004.html.

YU Y X, LI X. Research on text annotation method of ancient works from the perspective of digital humanism - A case study on MARKUS[J/OL]. Big data research: 1-11[2022-05-16]. http://hfghi60aabc7d15084b00h060wbob5fqnx6n66.fhcz.libproxy.ruc.edu.cn/kcms/detail/10.1321.G2.20220401.1032.004.html.

[25] 北京大学信息管理系. 数字人文视角下的典籍传承与学科融合论坛在北京大学举行[EB/OL]. (2019-10-16)[2022-05-15]. http://www.im.pku.edu.cn/xwgg/xgxw/323111.htm.

Department of information management, Peking university. The forum on the inheritance of classics and the integration of disciplines from the perspective of digital humanities was held at Peking university[EB/OL]. (2019-10-16)[2022-05-15]. http://www.im.pku.edu.cn/xwgg/xgxw/323111.htm.

[26] 索传军. 知识转移视角下的学术论文老化与创新研究[J]. 图书情报工作, 2014, 58(5): 5-12.

SUO C J. Study on obsolescence and innovation of academic papers from perspective of knowledge transfer[J]. Library and information service, 2014, 58(5): 5-12.

[27] 夏翠娟, 张磊. 关联数据在家谱数字人文服务中的应用[J]. 图书馆杂志, 2016(10):26-34.

XIA C J, ZHANG L. The application of linked data in shanghai library's service of genealogy digital humanities[J]. Library journal, 2016(10): 26-34.

[28] 王丽丽, 朱小梅. 古籍钤印元数据著录规范设计与应用研究[J]. 图书馆, 2020(1):106-111.

WANG L L, ZHU X M. Design and application of the seals metadata description criteria in ancient books[J]. Library, 2020(1): 106-

111.

[29] 欧阳剑. 面向数字人文研究的大规模古籍文本可视化分析与挖掘[J]. 中国图书馆学报, 2016(2): 66-80.
OUYANG J. Visual analysis and exploration of ancient texts for digital humanities research[J]. Journal of library science in China, 2016(2): 66-80.

Knowledge Correlation of Chinese Ancient Books from the Perspective of Digital Humanities

WANG Lili[1,2,3], ZHANG Ning[1,2*]

(1. Renmin University of China Libraries, Beijing 100872; 2. Institute of Renmin University Ancient Books Collocation and Studies, Beijing 100872; 3. Research Center for Digital Humanities of RUC, Beijing 100872)

Abstract: [Purpose/Significance] Chinese ancient books are one of the most important carriers to condense the essence of Chinese traditional culture. However, scholars have found that the database of Chinese ancient books only provides book metadata, which can only meet the basic retrieval and use needs of users, but does not dig and utilize Chinese ancient books in depth from the perspective of knowledge organization. The research in digital humanities is in the ascendant, providing an important modern scientific and technological means for the realization of knowledge correlation of Chinese ancient books. The research on the content mining and knowledge correlation of Chinese ancient books will help to improve the development level and service quality of Chinese ancient books, and promote the creative transformation and innovative development of Chinese excellent traditional culture. [Method/Process] Through literature review and online investigation, this paper summarizes the research of Chinese ancient books' knowledge correlation from the perspective of digital humanities. Based on the research, this paper discusses the differences between traditional database and digital humanistic platforms of Chinese ancient books. First, the organization mechanism is different, and the latter has more diverse and in-depth functions. Second, the scope of collection is expanded. The latter includes not only the original documents, but also the analytical texts and extended documents of ancient books, etc. In the past, most database platforms were "digital" platforms; The digital humanistic platform of Chinese ancient books is a platform of "data", which has been upgraded from "storage" to "analysis" and "correlation". Then, the key links of Chinese ancient books' knowledge correlation are discussed, and the barriers of knowledge correlation are analyzed by examples. [Results/Conclusions] From the perspective of digital humanities, collation of Chinese ancient books has developed from document units to knowledge units by using the data analysis technology, machine learning, visualization technology, etc. The realization of knowledge correlation of Chinese ancient books starts from document organization, based on ancient book data, and its essence is knowledge organization. Knowledge correlation of Chinese ancient books can be used for local chronicles and genealogy mining, seal knowledge network discovery, research context and communication analysis of scholars in specific periods and regions. However, the current knowledge correlation of ancient books still faces barriers such as the difficulty of knowledge element processing, high technical requirements and the lack of integrated platforms in its implementation. Limited by the research conditions, this paper only discusses the research and application, key links and barriers of ancient books' knowledge correlation from a theoretical point of view. In the future work, we will take the opportunity of optimizing and upgrading our ancient books' knowledge platform, to further study the metadata description specification, text conversion and content mining of Chinese ancient books, data exchange and update mechanism among different platforms.

Keywords: collation of ancient books; digital humanity; knowledge correlation; knowledge units